CATVis: Context-Aware Thought Visualization

Tariq Mehmood[1*][0009−0007−2743−0616], Hamza Ahmad[2*][0009−0005−5111−4547], Muhammad Haroon Shakeel[3][0000−0001−6237−3388], and Murtaza Taj[1][0000−0003−2353−4462]

[1] Lahore University of Management Sciences, Lahore, Pakistan
{23030009,murtaza.taj}@lums.edu.pk
http://cvlab.lums.edu.pk
[2] Forman Christian College, University, Lahore, Pakistan
241548068@formanite.fccollege.edu.pk
[3] Arbisoft, Lahore, Pakistan
mhfateen@gmail.com

Abstract. EEG-based brain-computer interfaces (BCIs) have shown promise in various applications, such as motor imagery and cognitive state monitoring. However, decoding visual representations from EEG signals remains a significant challenge due to their complex and noisy nature. We thus propose a novel 5-stage framework for decoding visual representations from EEG signals: (1) an EEG encoder for concept classification, (2) cross-modal alignment of EEG and text embeddings in CLIP feature space, (3) caption refinement via re-ranking, (4) weighted interpolation of concept and caption embeddings for richer semantics, and (5) image generation using a pre-trained Stable Diffusion model. We enable context-aware EEG-to-image generation through cross-modal alignment and re-ranking. Experimental results demonstrate that our method generates high-quality images aligned with visual stimuli, outperforming SOTA approaches by 27.08% in Classification Accuracy, 15.21% in Generation Accuracy and reducing Fréchet Inception Distance by 36.61%, indicating superior semantic alignment and image quality. The code is available at https://github.com/CVLABLUMS/CATVis.

Keywords: Brain-Computer Interfaces (BCIs) · EEG Decoding · Visual Reconstruction · Cross-Modal Alignment

1 Introduction

Brain-computer interfaces (BCIs) decode neural signals to enable communication, particularly for individuals with severe neuromotor impairments. Electroencephalography (EEG), a non-invasive and affordable technique, is widely used in BCIs due to its high temporal resolution, despite challenges such as low spatial resolution [26], low signal-to-noise ratio [3], and artifact contamination [12]. Decoding visual representations from EEG remains difficult, as traditional feature-based approaches struggled with limited datasets, while recent deep learning advances have improved reconstruction fidelity [8,25].

---

* These authors contributed equally to this work and are co-corresponding authors.

Existing approaches like ThoughtViz [24] and Brain2Image [11] primarily focus on visualizing the conceptual information captured by neural signals, often overlooking the contextual elements, such as color, setting, or interaction. By neglecting these finer details, current methods provide only a coarse approximation of the visual stimulus. To address this gap, we propose a framework that not only improves the concept recognition but also captures the finer details by incorporating textual descriptions as an intermediary representation.

Our framework employs the EEG–image pairs dataset from [14, 22], in which each EEG sample is paired with a corresponding visual stimulus. We begin by generating a caption for each stimulus to obtain a ground-truth description that serves as the basis for alignment in the CLIP [16] feature space. Next, our encoder classifies the EEG signals into conceptual categories, and we interpolate the concept and caption embeddings to form a semantically rich conditioning vector for Stable Diffusion (SD), enabling the generation of high-fidelity images.

Our end-to-end approach leverages state-of-the-art techniques in natural language processing and computer vision to create a complete pipeline from neural activity to visual output. Our contributions can be summarized as follows:

- Unlike prior work [11, 24], which focused on conceptual decoding (class labels), we enable context-aware (descriptive information) EEG-to-image generation through cross-modal alignment and re-ranking, boosting generation accuracy by 15.21% over baselines.
- Unlike resource-intensive self-supervised methods [2, 7], we achieve a 27.08% higher classification accuracy using a supervised EEG Conformer [21], demonstrating that robust EEG representations can be learned without extensive pretraining.
- We modulate the emphasis on concept vs. context. This increases the semantic fidelity of reconstructed images, aligning them more closely with the true visual content.

2   Related Work

Decoding visual information from brain activity and reconstructing images has gained significant attention in recent research. Initial work in this domain leveraged the discriminative power of GANs to directly generate realistic images from brain activity [11, 13, 20, 24]. Brain2Image [11] used an LSTM to classify EEG into 40 ImageNet classes, conditioning VAE and GAN generation with discriminative features. ThoughtViz [24] built on this with 1D/2D-CNNs to classify EEG into object classes, driving GAN synthesis. Yet, both [11, 24] rely on class-discriminative features, limiting their capacity to capture contextual details beyond basic conceptual decoding, thus reducing their utility for detailed visual reconstruction.

Diffusion models, inspired by non-equilibrium thermodynamics, have emerged as an effective generative modeling approach [17]. DreamDiffusion [2] fine-tunes Stable Diffusion [17] by conditioning on robust EEG features obtained through

a Masked Autoencoder approach. These features are further aligned with CLIP image embeddings to improve cross-modal consistency, enabling high-quality image generation. However, DreamDiffusion relies on computationally intensive self-supervised pretraining, which limits its scalability and accessibility, particularly for resource-constrained settings.

In contrast, BrainVis [7] removes the need for external data by training a robust EEG encoder with Latent Masked Modeling, leveraging both time-domain and frequency-domain features while incorporating reconstruction and classification objectives. These features are aligned with interpolated semantic embeddings (coarse-grained labels and fine-grained captions). The aligned embeddings are used as conditioning for cascaded diffusion models, resulting in better semantic accuracy of generated images. However, BrainVis' complex cascaded diffusion architecture increases computational overhead.

3 Methodology

Our current methodology comprises five main components: (1) Conformer-Based EEG Encoder for concept classification; (2) Cross-modal alignment in CLIP feature space between EEG and caption embeddings; (3) Caption refinement through re-ranking of retrieved embeddings from CLIP feature space; (4) Weighted interpolation of concept and caption for semantically rich embedding; (5) Image generation using pretrained Stable Diffusion.
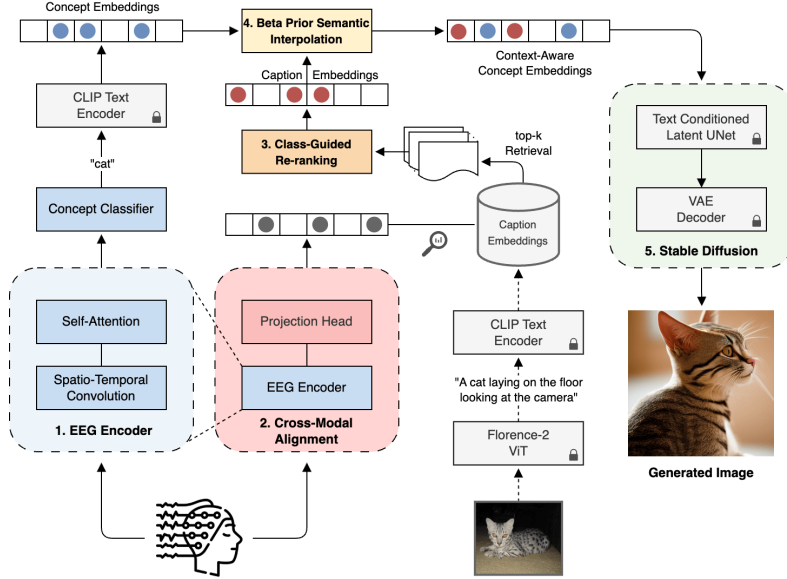

Fig. 1. Overview of our proposed five-stage framework for reconstructing images from EEG signals. The numbered modules correspond to the key steps in the methodology.

3.1 EEG Encoder for Concept Classification

Unlike self-supervised approaches [2,7] that rely on self-supervised reconstruction objectives, we adopt a supervised EEG Conformer [21] encoder, trained from scratch on our dataset. The architecture leverages spatio-temporal convolutions and a stacked Transformer encoder to capture EEG signals' local and long-term dependencies, with a fully connected classification head to predict class labels.

3.2 Cross-modal Alignment

Our cross-modal alignment framework learns to project EEG signals into CLIP's joint embedding space through contrastive learning. We adapt the EEG Conformer architecture by replacing its final classification layer with a linear projection to 768 dimensions followed by L2-normalization. The caption embeddings are obtained via CLIP's pretrained ViT-L/14 text encoder. We optimize the following symmetric contrastive loss:

$$\mathcal{L} = \frac{1}{2}(\mathcal{L}_{\text{EEG}\rightarrow\text{Text}} + \mathcal{L}_{\text{Text}\rightarrow\text{EEG}}) \tag{1}$$

To be more explicit, we have two directional terms. For the EEG→Text direction:

$$\mathcal{L}_{\text{EEG}\rightarrow\text{Text}} \;=\; -\sum_{i=1}^{B} \log\left(\frac{\exp\!\big(\mathbf{e}_i^{\text{EEG}} \cdot \mathbf{e}_i^{\text{text}} \,/\, \tau\big)}{\sum_{j=1}^{B} \exp\!\big(\mathbf{e}_i^{\text{EEG}} \cdot \mathbf{e}_j^{\text{text}} \,/\, \tau\big)}\right), \tag{2}$$

where $\mathbf{e}_i^{\text{EEG}}$ is the normalized embedding for the EEG sample i, $\mathbf{e}_j^{\text{text}}$ is the embedding for the text sample j, B is the mini-batch size, and τ is the temperature parameter. Similarly, for the Text→EEG direction:

$$\mathcal{L}_{\text{Text}\rightarrow\text{EEG}} \;=\; -\sum_{i=1}^{B} \log\left(\frac{\exp\!\big(\mathbf{e}_i^{\text{text}} \cdot \mathbf{e}_i^{\text{EEG}} \,/\, \tau\big)}{\sum_{j=1}^{B} \exp\!\big(\mathbf{e}_i^{\text{text}} \cdot \mathbf{e}_j^{\text{EEG}} \,/\, \tau\big)}\right). \tag{3}$$

In this setup, each EEG embedding $\mathbf{e}_i^{\text{EEG}}$ is "pulled closer" to its corresponding text embedding $\mathbf{e}_i^{\text{text}}$ and pushed away from non-matching captions in the batch. Figure 2 demonstrates this alignment process.

3.3 Caption Refinement

Through cross-modal alignment, we align EEG representations with CLIP embeddings, which contain fine-grained semantic information. However, because EEG signals reflect only the subject's perception rather than every detail of the captioned image, they often fail to capture finer contextual cues. As a result, the neural response typically centers on the main concept and some contextual elements, leading to alignment inconsistencies when EEG embeddings are compared with CLIP text embeddings containing richer semantic information.

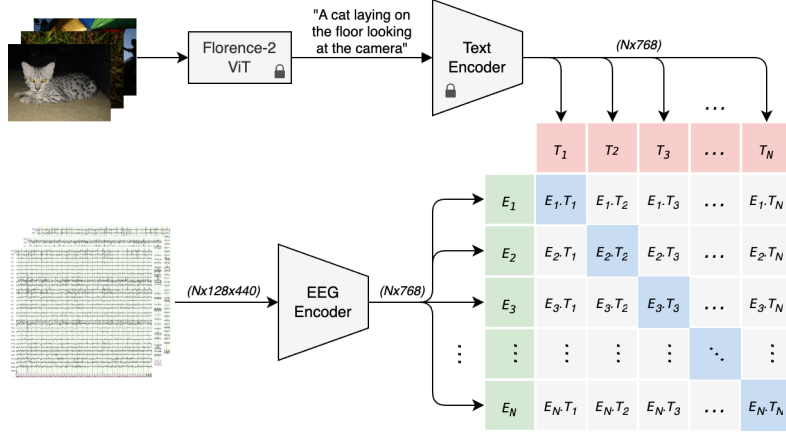To address this limitation, we employ a two-stage retrieval process:

Fig. 2. Cross-modal alignment of EEG and text embeddings via contrastive loss. We project EEG signals and captions into a joint representation space, where positive pairs (blue cells) are brought closer, and negative pairs (gray cells) are pushed apart.

1. **Initial Retrieval.** Given an EEG embedding $\mathbf{e}^{\text{EEG}} \in \mathcal{Q}$ (the *query*) and a collection of text embeddings $\mathcal{X} = \{\mathbf{e}_1^{\text{text}}, \ldots, \mathbf{e}_n^{\text{text}}\}$ of n caption embeddings belonging to the entire corpus, we retrieve the top-k most similar captions in the CLIP feature space as:

$$\underset{x \in \mathcal{X}}{\arg \max^{(k)}} \cos\left(\mathbf{e}^{\text{EEG}}, \mathbf{e}^{\text{text}}\right), \tag{4}$$

   This operation returns a subset of $\mathcal{X}$ that best match the neural response.
2. **Class-Guided Re-ranking.** These top-k candidates are then re-ranked according to their similarity to the predicted CLIP-encoded class label obtained from our concept classifier, ensuring that the final caption aligns with the primary object most likely perceived by the subject. This re-ranking step yields 6.99% higher Generation Accuracy (GA), 6.30% higher Inception Score (IS), and 3.43% lower Fréchet Inception Distance (FID) versus without re-ranking (see Table 2).

3.4   Beta Prior Semantic Interpolation

To balance object-focused conceptual and fine-grained descriptive information, we interpolate the predicted concept embedding from our EEG classifier with the re-ranked caption embedding. Let $\mathbf{e}^{\text{class}} \in \mathbb{R}^{768}$ be the CLIP-encoded class label predicted by the EEG Conformer, and $\mathbf{e}^{\text{text}} \in \mathbb{R}^{768}$ the top-ranked CLIP caption embedding. We sample $\lambda \sim \text{Beta}(\alpha, \beta)$ and compute:

$$\mathbf{z} = \lambda \mathbf{e}^{\text{class}} + (1 - \lambda)\mathbf{e}^{\text{text}}, \tag{5}$$

where $\mathbf{z}$ conditions the diffusion process.

3.5   Stable Diffusion

We use a pre-trained Stable Diffusion model [17] to transform the EEG-derived conditioning vector $\mathbf{z}$ into photorealistic images. Specifically, we condition the denoising process on $\mathbf{z}$ by integrating it into the cross-attention mechanism of Stable Diffusion's UNet, guiding the reverse diffusion process to reconstruct the image latent. Finally, the VAE decoder converts the latent representation into the pixel space, yielding the generated image. By leveraging Stable Diffusion's generative capabilities with EEG-driven embeddings, our approach aims to generate images that capture both the primary object class and finer contextual details inferred from neural activity.

4   Experiments

4.1   Dataset

We use the corrected EEG-ImageNet dataset from [14], which addresses preprocessing flaws in the original release [22], such as the inclusion of a DC component that introduced EEG drift and artificially boosted performance. This dataset has been the subject of methodological discussion: [1] highlights potential experimental confounds, including block-level stimulus leakage and classifier bias. In response, [15] provide a detailed rebuttal, citing cognitive neuroscience literature in support of block designs and contesting the validity of the critique's single-subject, low-frequency evaluation setup. While the EEG-ImageNet dataset stands as valid and citable, newer alternatives such as [9] also offer a valuable direction for future work.

The EEG-ImageNet dataset [14] features EEG signals from 6 subjects exposed to 2000 ImageNet images (40 concepts, 50 images each) via Rapid Serial Visual Presentation, recorded with a 128-electrode headset. Each trial yields $12{,}000$ EEG-image pairs, with signals (128 channels, 440 timepoints) preprocessed to retain the $20 - 460$ ms interval and band-pass filtered at $55 - 95$ Hz for cognitive relevance [14]. We adopt the provided train/validation/test splits. Ground-truth captions are generated using Microsoft Florence-2 Large [27] to pair EEG signals with image descriptions for cross-modal alignment.

4.2   Implementation Details

We implement our framework in PyTorch (Python 3.10), using the EEG Conformer [21] as the encoder backbone, modified to output 768-dimensional embeddings for CLIP alignment. The temporal convolution uses $k = 40$ kernels of size $(1, 25)$, while the spatial convolution applies the same number of kernels with a size of $(ch, 1)$, where $ch = 128$. The encoder is trained for 120 epochs with Adam ($\eta = 0.0001$) on cross-entropy loss. The cross-modal alignment module is trained separately for 100 epochs, minimizing InfoNCE loss ($\tau = 0.07$) with random sampling for mini-batches. Interpolation coefficients are sampled from a Beta distribution ($\alpha = 10, \beta = 10$). Image generation uses pre-trained Stable Diffusion v1-5 [17], conditioned via CLIP ViT-L/14 embeddings.

4.3   Evaluation Metrics

EEG Classification Accuracy (CA) CA measures the EEG encoder's ability to extract discriminative latent features from EEG signals, with higher values indicating better class prediction accuracy.

N-way Top-K Classification Accuracy of Generation (GA) Following [2, 4, 7], GA evaluates semantic correctness of generated images using an N-way top-K classification task. We use a pre-trained ImageNet1K classifier [5] to compare ground-truth and generated images, selecting N-1 random classes plus the ground-truth class, and checking if the top-K prediction matches the ground truth over 50 trials. We set $N = 50$ and $K = 1$, consistent with prior evaluations.

Inception Score (IS) IS [18] assesses image quality and diversity using a pre-trained Inception-v3 classifier [23]. It measures classifier confidence and class distribution diversity, where higher scores reflect better quality and variety.

Fréchet Inception Distance (FID) FID [10] evaluates generation quality by computing the feature distribution distance between ground-truth and generated images using Inception-v3 [23]. Lower FID values indicate better alignment.

5   Results & Discussion

Our framework demonstrates significant advancements in decoding visual representations from EEG signals, as evidenced by both quantitative metrics and qualitative outputs. Table 1 shows that our supervised EEG Conformer achieves a Top-1 CA of 61.09%, surpassing BrainVis [7] by 27.08%, highlighting the encoder's ability to extract robust, discriminative features without resource-intensive pretraining. For image generation, our method achieves a GA of 56.78% for Subject 4 (Table 4) and 52.65% on average (Table 3), outperforming DreamDiffusion by 23.91% (Subject 4) and BrainVis by 15.21% (average), alongside an IS of 37.49 and a FID of 80.29, reflecting a 36.61% reduction in FID compared to BrainVis. These results underscore improved semantic alignment and image quality, as visually confirmed in Figure 3, where generated images closely resemble their ground-truth stimuli in both conceptual and contextual details. Collectively, these findings validate the efficacy of our five-stage approach.

5.1   Ablation Studies

We ablated our EEG Conformer and interpolation strategy to assess their roles. Table 5 shows removing the Transformer reduces Top-1 CA by 5.76%, as it captures global temporal dependencies across EEG channels, vital for long-range neural patterns. Excluding the convolutional block drops CA by 45.77%, since it extracts local spatio-temporal features essential for stimulus-specific responses.
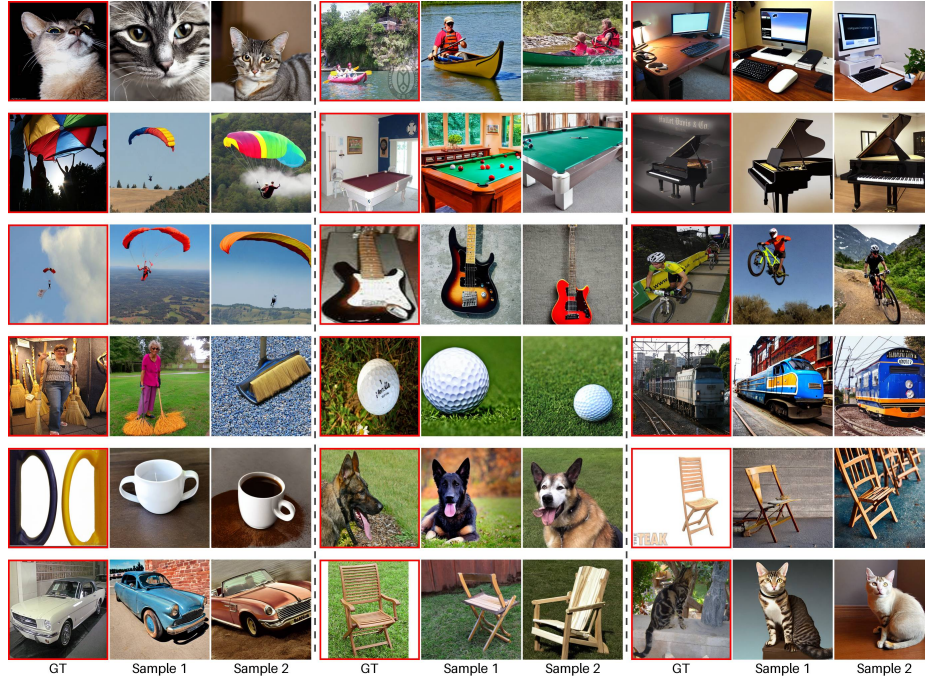
Fig. 3. Images generated from EEG signals using our framework. Red boxes show visual stimuli, followed by two generated samples. First 15 images (i.e., the first 5 rows) are the same visual stimuli used in Fig. 6 of [7] to allow qualitative comparison.

Table 6 tests interpolation weights: a concept-heavy weight (0.75) yields the highest GA (61.48%) by emphasizing class-specific accuracy from the EEG classifier, while a caption-heavy weight (0.75) lowers GA (53.93%) due to potential noise in descriptive details. The Beta(10,10) scheme balances both, achieving the best IS (35.79) by enhancing diversity via captions. These results validate the synergy of encoder components and the interpolation's role in balancing conceptual and contextual fidelity for Stable Diffusion conditioning.

6   Conclusion

Our five-stage framework advances EEG-based visual reconstruction, achieving a 27.08% higher CA and 15.21% improved GA over state-of-the-art methods, with a 36.61% FID reduction. By integrating supervised EEG encoding, CLIP alignment, caption refinement, and interpolation, we capture both conceptual and contextual details, outperforming prior approaches without extensive pre-training. This work offers a scalable solution for BCI applications, with future potential in refining cross-modal alignment and exploring real-time visualization.

Table 1. Comparison of EEG classification accuracy (%) with existing methods.

Method	Top1↑	Top3 ↑	Top5↑	F1 ↑
Brain2Image	16.48	37.90	55.24	0.16
KD-STFT [6]	41.20	75.30	87.82	0.40
BrainVis	47.66	79.23	91.20	0.43
CATVis	**61.09**	**93.84**	**98.15**	**0.61**

Table 2. Effect of caption re-ranking on generation performance.

Method	GA (%↑)	IS ↑	FID ↓
w/o Re-ranking	53.07	32.46	86.07
w/ Re-ranking	**56.78**	**34.51**	**83.12**

Table 3. Comparison with image reconstruction baselines on all six subjects.

Method	GA (%↑)	IS ↑	FID ↓
Brain2Image	-	5.07	-
ESG-ADA [19]	-	10.82	174.13
BrainVis	45.46	31.00	126.66
CATVis	**52.65**	**37.50**	**80.30**

Table 4. Comparison with image reconstruction baselines on Subject 4.

Method	GA (%↑)	IS ↑	FID ↓
DreamDiffusion	45.80	-	-
BrainVis	49.27	31.53	121.02
CATVis	**56.78**	**34.51**	**83.13**

Table 5. Classification accuracy (%) of EEG Encoder w/wo CNN and w/wo Transformer Encoder (TE)

CNN	TE	Top1 ↑	Top3 ↑	Top5 ↑	F1 ↑
-	✓	15.32	34.50	48.27	0.16
✓	-	55.33	90.84	97.70	0.55
✓	✓	**61.09**	**93.84**	**98.15**	**0.61**

Table 6. Ablation on concept–caption interpolation with weights $\omega_{concept}, \omega_{caption}$.

$\omega_{concept}$	$\omega_{caption}$	GA (%↑)	IS ↑	FID ↓
0.75	0.25	**61.48**	34.34	**82.33**
0.25	0.75	53.93	30.50	83.30
0.5	0.5	56.79	34.50	86.66
Beta(10,10)		56.86	**35.79**	84.13

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahmed, H., Wilbur, R.B., Bharadwaj, H.M., Siskind, J.M.: Confounds in the data—comments on "Decoding brain representations by multimodal learning of neural activity and visual features". IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(12), 9217–9220 (2021)
2. Bai, Y., Wang, X., Cao, Y.p., Ge, Y., Yuan, C., Shan, Y.: DreamDiffusion: Generating high-quality images from brain EEG signals. arXiv preprint arXiv:2306.16934 (2023)
3. Chaddad, A., Wu, Y., Kateb, R., Bouridane, A.: Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques. Sensors **23**(14), 6434 (2023)
4. Chen, Z., Qing, J., Xiang, T., Yue, W.L., Zhou, J.H.: Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding. In: IEEE Int. Conf. on Computer Vision and Pattern Recognition. pp. 22710–22720. IEEE (2023)
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: Int. Conf. on Learning Representations (2021)

6. Ferrante, M., Boccato, T., Bargione, S., Toschi, N.: Decoding EEG signals of visual brain representations with a CLIP based knowledge distillation. In: ICLR Workshop on Learning from Time Series For Health (2024)

7. Fu, H., Wang, H., Chin, J.J., Shen, Z.: BrainVis: Exploring the bridge between brain and visual signals via image reconstruction. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing. pp. 1–5. IEEE (2025)

8. Gong, S., Xing, K., Cichocki, A., Li, J.: Deep learning in EEG: Advance of the last ten-year critical period. IEEE Transactions on Cognitive and Developmental Systems **14**(2), 348–365 (2022)

9. Guenther, S., Kosmyna, N., Maes, P.: Image classification and reconstruction from low-density EEG. Scientific Reports **14**(1), 16436 (2024)

10. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. Advances in Neural Information Processing Systems **30** (2017)

11. Kavasidis, I., Palazzo, S., Spampinato, C., Giordano, D., Shah, M.: Brain2Image: Converting brain signals into images. In: ACM Int. Conf. on Multimedia. pp. 1809–1817. ACM (2017)

12. Mannan, M.M.N., Kamran, M.A., Jeong, M.Y.: Identification and removal of physiological artifacts from electroencephalogram signals: A review. IEEE Access **6**, 30630–30652 (2018)

13. Mishra, R., Sharma, K., Jha, R.R., Bhavsar, A.: NeuroGAN: Image reconstruction from EEG signals via an attention-based GAN. Neural Computing and Applications **35**(12), 9181–9192 (2023)

14. Palazzo, S., Spampinato, C., Kavasidis, I., Giordano, D., Schmidt, J., Shah, M.: Decoding brain representations by multimodal learning of neural activity and visual features. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(11), 3833–3849 (2020)

15. Palazzo, S., Spampinato, C., Kavasidis, I., Giordano, D., Schmidt, J., Shah, M.: Rebuttal to "Comments on 'Decoding brain representations by multimodal learning of neural activity and visual features'". IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(12), 11540–11542 (2024)

16. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: Int. Conf. on Machine Learning. pp. 8748–8763. PMLR (2021)

17. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: IEEE Int. Conf. on Computer Vision and Pattern Recognition. pp. 10684–10695. IEEE (2022)

18. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training GANs. In: Advances in Neural Information Processing Systems. vol. 29 (2016)

19. Singh, P., Dalal, D., Vashishtha, G., Miyapuram, K., Raman, S.: Learning robust deep visual representations from EEG brain recordings. In: IEEE Winter Conf. on Applications of Computer Vision. pp. 7553–7562. IEEE (2024)

20. Singh, P., Pandey, P., Miyapuram, K., Raman, S.: EEG2IMAGE: Image reconstruction from EEG brain signals. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing. pp. 1–5. IEEE (2023)

21. Song, Y., Zheng, Q., Liu, B., Gao, X.: EEG Conformer: Convolutional transformer for EEG decoding and visualization. IEEE Transactions on Neural Systems and Rehabilitation Engineering **31**, 710–719 (2023)

22. Spampinato, C., Palazzo, S., Kavasidis, I., Giordano, D., Souly, N., Shah, M.: Deep learning human mind for automated visual classification. In: IEEE Int. Conf. on Computer Vision and Pattern Recognition. pp. 6809–6817. IEEE (2017)
23. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: IEEE Int. Conf. on Computer Vision and Pattern Recognition. IEEE (2016)
24. Tirupattur, P., Rawat, Y.S., Spampinato, C., Shah, M.: ThoughtViz: Visualizing human thoughts using generative adversarial network. In: ACM Int. Conf. on Multimedia. pp. 950–958. ACM (2018)
25. Vallabhaneni, R.B., Sharma, P., Kumar, V., Kulshreshtha, V., Reddy, K.J., Kumar, S.S., Kumar, V.S., Bitra, S.K.: Deep learning algorithms in EEG signal decoding application: A review. IEEE Access **9**, 125778–125794 (2021)
26. Värbu, K., Muhammad, N., Muhammad, Y.: Past, present, and future of EEG-based BCI applications. Sensors **22**(9), 3331 (2022)
27. Xiao, B., Wu, H., Xu, W., Dai, X., Hu, H., Lu, Y., Zeng, M., Liu, C., Yuan, L.: Florence-2: Advancing a unified representation for a variety of vision tasks. In: IEEE Int. Conf. on Computer Vision and Pattern Recognition. pp. 4818–4829. IEEE (2024)